# Proving Ground: 2,048 Intel® Xeon™ Processors Power Los Alamos' "Science Appliance" with 9.2 Tflops

## SOLUTION SUMMARY

| | |
|---|---|
| Challenge | Los Alamos National Laboratory, one of the nation's leading research labs, needed a large-scale cluster that would provide a proving ground for new computer science technologies and offer additional compute power for its Grand Challenge science initiatives. To advance the scalability of large-scale clusters, Los Alamos wanted the largest open source cluster it could afford that would also support LinuxBIOS*, an open source BIOS alternative. |
| Solution | Los Alamos is deploying an innovative, diskless cluster that harnesses the power of 2,048 Intel® Xeon™ processors to pack theoretical peak performance of 9.2 Tflops into a footprint of just over 450 square feet. Using LinuxBIOS and other software, Los Alamos expects to be able to boot the massive system in under 30 seconds. |
| Business value | Los Alamos' new "Science Appliance" significantly expands the computational power available to nonclassified research across the lab's mission-space, empowering scientists to achieve better results faster and advance the state of knowledge in their respective fields. This work advances the feasibility of using clusters for ASCI-class systems in nuclear weapons applications and others and will help create next-generation clusters that are more dense, powerful, manageable and reliable. |
| High performance cluster | 1,024 dual-processor server nodes fully populated with 2,048 Intel Xeon processors at 2.40 GHz and Intel® E7500 chipsets. Nodes are based on Linux NetworX Evolocity* II platform and Supermicro motherboards. The system has 2.1 terabytes of aggregate memory. |
| Operating system | Customized version of Red Hat Linux* incorporating LinuxBIOS |
| Interconnect | Myricom Myrinet* |
| Cluster management | Linux NetworX ClusterWorX*, ICE Box* 3.0 appliance, Clustermatic and BProc open source software |
| System integration | Los Alamos, Linux NetworX |

## Challenge

### NEVER ENOUGH

From bioinformatics to nuclear weapons performance and safety, Los Alamos National Laboratory uses state-of-the-art computers to conduct some of the most sophisticated research in the world. For its scientists, as for technologists in such fields as aerospace design, bioengineering and materials science, there's no such thing as enough computing power. True, each new generation of more capable systems enables technologists to more closely model physical reality, integrate more physics into their simulations and examine more complex problems in greater detail. But each set of answers opens the door to new questions that demand still

**"We have built our last proprietary system. We are fully committed to open standards, open architectures and open source."**

**Ron Minnich**
Cluster Research Team Leader
Advanced Computing Lab
Los Alamos National Laboratory

more computational horsepower. And the challenges that face the nation, ranging from counterterrorism to global warming, continue to increase the demands on the country's high performance computing (HPC) infrastructure.

"In the sciences and engineering, there's always been an insatiable appetite for computing horsepower," explains Bill Feiereisen, leader of Los Alamos' Computer and Computational Sciences Division. "Our traditional, science-based disciplines will eat up any horsepower we can provide. Now it's becoming clear that the data processing and communications needs of homeland security and the war on terrorism will also demand powerful but economical computing."

Los Alamos is betting on off-the-shelf microprocessors and open source operating systems to deliver the hundreds of teraflops of sustained performance that Los Alamos and other national labs will need later in the decade. To make those supersized clusters a practical reality, Los Alamos is installing an enormous Intel® Xeon™ processor-based cluster it will use to develop, scale and test technologies that will improve the reliability and manageability of next-generation clusters. The new cluster, which lab technologists have dubbed the Science Appliance, will also support research in a variety of non-classified Grand Challenge applications. The cluster is being purchased with money from the National Nuclear Security Administration's Advanced Simulation and Computing program.

"Future supercomputers must be cost-effective, efficient and easy to enhance and scale," Feiereisen says. "Scalable supercomputing systems that run proprietary operating systems clearly are a thing of the past. Instead of buying a complete proprietary computing system, we are looking toward a future in which a robust set of integrated, open source software tools enables us to assemble a truly scalable supercomputer from components that best meet our needs."

### BEYOND PROPRIETARY

Los Alamos has a long tradition of leadership in research critical to the nation's well-being. Established in 1943 as part of the top-secret Manhattan Project, Los Alamos is located on a 43-square-mile site in the mountains northwest of Santa Fe, New Mexico. The lab is part of the U.S. Department of Energy's National Nuclear Security Administration and is operated by the University of California. Since 1978, Los Alamos has earned 79 R&D 100 awards, given by *R&D Magazine* to recognize the year's 100 most significant advances in science and technology. The lab has more than 12,000 full-time equivalent employees and an FY2002 budget of $1.796 billion.

Like computational science researchers across the public and private sectors, those at Los Alamos National Lab rely on advances in high performance computing to enable new discoveries, answer critical questions and identify promising areas of inquiry. Shifting from proprietary architectures to an open architecture and open source operating system is a win for those scientists on several fronts. An open architecture enables Los Alamos to take advantage of the rapid performance roadmap and cost savings of off-the-shelf microprocessors. It also allows the lab to assemble best-of-breed solutions.

"Our applications workload is somewhat unique because we have integrated, multi-scale physics that places significant performance demands on everything from memory subsystems to interconnects to I/O subsystems," explains deputy division leader Stephen R. Lee. "An open environment supported by open source software enables us and in fact the HPC community to select the computational components—interconnect, processor and so forth—that are best suited to our applications."

The shift also enables the lab to take advantage of the wealth of resources available in the open source community. "With Linux*, the team of bug fixers is global," says Ron Minnich, cluster research team leader at Los Alamos' prestigious Advanced Computing Lab. "If I run into a problem, I can post a message or e-mail someone about it and have a fix within a day. The scope of expertise is amazing."

The contribution flow runs both directions, simplifying the important work of technology transfer and making it easy for Los Alamos to share its advances with a broader community. "Adopting an open source operating system and software suite invites immediate collaboration in an environment that many people are familiar with," Feiereisen says. "It gives us a repository for HPC developments that benefit the entire computing community."

Charles Strauss, a computational biology researcher in Los Alamos' Biosciences Division, is one of many lab scientists who will benefit from the new cluster's high performance. Strauss's team is developing advanced algorithms for amino acid sequence analysis—a field that is growing in significance since the mapping of the human genome.

"Large cluster-computing type applications are essential to my work," Strauss explains. "Reducing the time it takes to get a result is very important because tuning all the parameters of the algorithm becomes increasingly complex. You have to perform more and more test runs each time. If you look at our ability to predict structures, it's pretty much tracked Moore's Law in what we can do."

## Solution

### INNOVATING ON INTEL® ARCHITECTURE

As an HPC innovator, Los Alamos gets another important benefit from adopting open solutions: Reducing the thrashing that occurs when adopting a new platform

> **“The bandwidth between the processor and main memory is hands-down better than anything else on the market....The new Intel® compilers are a quantum leap forward and do a great job of helping you extract maximum performance out of the hardware. It all adds up to a very high throughput system for high performance computing.”**
>
> **Joshua Harr**
> Chief Technology Officer
> Linux NetworX

architecture. “Our research creates a tremendous amount of intellectual property that goes in the trash when we move from one proprietary architecture to another,” says Minnich. “Open source lets us take that research with us when we move to a new platform. We have built our last proprietary system. We are fully committed to open standards, open architectures and open source.”

The Science Appliance brings to fruition key architectural innovations that Los Alamos has spearheaded over the past three years, as well as others developed in the lab and open source communities. For starters, the cluster is extremely dense and has a minimum of moving parts. “If you look at a typical HPC cluster, each node might have a couple serial ports, an Ethernet port, a high speed network, CD-ROM drive or hard disk and a full, installed operating system,” Minnich says. “Multiply that by 1,000 nodes or more, and you’re looking at a lot of points of failure. We wanted to eliminate as many of those as we could. By doing so, we improve reliability and reduce costs and power consumption.”

With the Science Appliance, system components are reduced to a bare—yet extremely powerful—minimum. Each node has two Intel Xeon processors running at 2.40 GHz, 2 GB of RAM, a flash memory device and several fans. Ethernet has been eliminated, and a high-speed Myricom Myrinet* network is used to handle management tasks and other communication between nodes and with the outside world, including loading the operating system. This eliminates the need for multiple ports and internal disk storage.

“You don’t need a permanent disk image of the operating system in the node. The nodes receive the operating system kernel from a mother node at startup and keep it resident during execution,” explains Josh Harr, chief technology officer of Salt Lake City-based Linux NetworX, the company that’s working with Los Alamos to integrate the system.

Loading the OS at system startup makes the cluster significantly easier to manage. “Cluster management is often compared to herding cats, because every disk has its own image,” Harr says. “If a node is out for repairs when you upgrade your system libraries, you could have random errors when you put it back and not realize where they’re coming from. The diskless approach ensures that everything is loaded identically and makes the system much more reliable.”

The approach also makes the Science Appliance an extremely dense HPC platform, with a footprint of roughly 450 square feet. Nodes are vertically stacked, 10 inside each subchassis, in 27 racks of 50 nodes each.

On the software side, the Science Appliance incorporates several technologies that Los Alamos had a hand in developing. LinuxBIOS* is key to making it possible to download the operating system and manage the system over the network. An open source BIOS that simplifies cluster management and boots nodes in seconds, LinuxBIOS replaces the normal BIOS bootstrap mechanism with a Linux kernel that can be booted from a cold start. Los Alamos uses the Beowulf Distributed Process Space (BProc) to provide a single system image of the entire cluster, hiding complexities and making the system easier to use and administer.

## CHOOSING THE INTEL® XEON™ PROCESSOR: COST, SOFTWARE AND PERFORMANCE

The Intel Xeon processor’s performance characteristics made it an outstanding match for Los Alamos’s science needs, according to Harr. “The Intel Xeon processor’s double-precision floating-point instructions are well suited to the needs of HPC users,” Harr says. “The bandwidth between the processor and main memory is hands-down better than anything else on the market. The Intel® E7500 chipset provides outstanding PCI performance and the Supermicro board has four separate PCI buses, so you’re not throttling down the Myrinet performance. The new Intel® compilers are a quantum leap forward and do a great job of helping you extract maximum performance out of the hardware. It all adds up to a very high throughput system for high performance computing.”

Linux NetworX had developed a robust LinuxBIOS port for the Intel E7500 chipset, and that was another item that brought the Intel Xeon processor to the top of Los Alamos’ rigorous bidding process. “LinuxBIOS was the first Go/No Go question,” says Minnich. “The fact that there was a supported LinuxBIOS coupled with the really attractive price/performance of the system made it a good choice for the Science Appliance.”

The system’s outstanding price/performance was another plus. “The goal was to assemble a cluster with as many nodes as possible to study scaling issues,” Feiereisen says. “We also wanted the nodes to have at least two processors in shared memory so we could also study SMP issues, race conditions and so forth.”

The quality of the Intel design, test and manufacturing processes are helping ensure a robust, reliable system for the lab. “Of the 400+ nodes that we’ve built and run through our 12-hour test system, we have had zero failures of the Intel® processor and an overall 98 percent acceptance rate,” says Dean Hutchings, chief operating officer at Linux NetworX. “That’s phenomenal for a system like this, and it speaks to the rigorous testing that Intel performs.”

After the sale, Intel proved itself highly responsive as a vendor and collaborator, according to Harr. That made a big difference to Los Alamos and Linux NetworX in working out the details of such a complex system. "When we ran into some down-and-dirty chipset issues, our sales engineer heard about the problem and solved it in record time," Harr says. "When a distributor had a problem that led to our getting some processors we couldn't use—right at a critical point in the project—Intel went through herculean efforts to fix the problem in time for us to meet our delivery deadlines."

The easy availability of critical information has also been a plus. "For the software work we do, we need extremely detailed information on the hardware," says Minnich. "We can download detailed chipset documentation right off the Intel Web site. Having that level of information makes a big difference."

### A NEW ERA

The power and cost-effectiveness of systems such as the Science Appliance give Los Alamos—and the nation—an important platform for next-generation research. "We are moving into an era where the combination of theory and methods, experimentation and simulation will create a predictive science capability for the nation," says Stephen Lee. "Our research projects require ever more powerful and complex simulation capability. This platform, together with the HPC Linux work that Ron and his team are doing, enable us to build simulation tools and applications best suited to the science we are examining, from stockpile stewardship to homeland defense and other important areas of investigation for the country."

### More Information

www.intel.com/eBusiness
www.ccs.lanl.gov
www.linuxbios.org
www.linuxnetworx.com

### LESSONS LEARNED

- **Performance, performance, performance.** From its instruction set to the compilers to the chipset's processor/memory bandwidth, the Intel® Xeon™ processor delivers outstanding performance for HPC clusters.
- **Let's get together.** The processor's and chipset's thermal and power characteristics are well suited to the demands of high performance clusters, helping Los Alamos and Linux NetworX create a dense and powerful HPC system.
- **Quality counts.** When you've got 2,000 of anything, even infrequent failures add up to a lot of downtime. Los Alamos is increasing uptime by eliminating unnecessary components and deploying a reliable, rigorously tested platform.
- **With an Intel®-based solution, you're backed by an entire ecosystem.** Whether Los Alamos needed high-speed interconnects or integration expertise, it could choose from a wide variety of vendors in an open marketplace. Intel itself backed them up with everything from in-depth chipset documentation to targeted problem-solving support.
- **Be open.** An open architecture and open source technologies reduce costs and make it easy to create systems tailored to the research focus. An open platform facilitates collaboration and information sharing and offers the opportunity to draw from a worldwide community.

*Solution provided by*
*Linux NetworX*

Linux NetworX
Powerful Cluster Technology

Intel works with the world's largest community of technology leaders and solution providers—from software and hardware to systems integration and services companies—that all are working with Intel® products, technologies and services with a common goal of providing better, more agile, cost-effective business solutions for you.

**Find out more about a business solution that is right for your company by contacting your Intel representative, or visit the Intel® Business Computing Web site at: intel.com/ebusiness or its industry solutions specific sites: intel.com/go/retail, intel.com/go/manufacturing, intel.com/go/digitalmedia, intel.com/go/finance, intel.com/go/telco, intel.com/go/hpc.**

intel®

Information in this document is provided in connection with Intel® products. Except as provided in Intel's terms and conditions of sale for such products, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY RELATING TO SALE AND/OR USE OF INTEL PRODUCTS, INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT, OR OTHER INTELLECTUAL PROPERTY RIGHT. Intel products are not intended for use in medical, life-saving, life-sustaining, critical control or safety systems, or in nuclear facility applications.

Intel may make changes to specifications, product descriptions, and plans at any time, without notice.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, reference www.intel.com/procs/perf/limits.htm, or call (U.S.) 1-800-628-8686 or 1-916-356-3104.

Intel, the Intel logo and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others.